

多模态数据融合与运动学特征驱动的手势动态检测系统研究

王乐萌^{1,2}

1. 中车科技创新(北京)有限公司, 中国·北京 100080
2. 中车工业研究院有限公司, 中国·北京 100071

摘要:针对手势动态检测中数据来源单一、规则适配依赖人工配置、运动过程识别稳定性不足等问题,提出一种多模态数据融合与运动学特征驱动的手势动态检测系统。以轨道交通司机作业合规检测为典型应用场景,系统融合路网信息、司机室视频与列车运行状态三类异构数据,利用大模型对路网信息进行语义解析生成结构化规则数据,结合当前运行状态动态匹配检测规则,并基于运动学特征对手势连续运动过程进行分析,实现对规范手势作业行为的合规性判定。该系统降低了人工配置规则的依赖,提高了复杂场景下规则适配的灵活性与合规检测的实时性。

关键词:多模态数据融合;大模型;运动学特征;手势检测;合规感知

Research on a Dynamic Gesture Detection System Driven by Multimodal Data Fusion and Kinematic Features

Wang Lemeng^{1,2}

1. CRRC Technology Innovation (Beijing) Co., Ltd., China Beijing 100080
2. CRRC ACADEMY CO., Ltd., China Beijing 100071

Abstract: To address issues such as single-source data dependency, reliance on manual rule configuration, and unstable recognition of gesture motion processes in dynamic gesture detection, this paper proposes a dynamic gesture detection system driven by multimodal data fusion and kinematic features. Taking compliance inspection of rail transit driver operations as a typical application scenario, the system integrates three heterogeneous data types—road network information, cab video, and train operating status. It employs a large language model to semantically parse road network information and generate structured rule data, dynamically matches detection rules based on the current operating status, and analyzes the continuous gesture motion process using kinematic features, thereby enabling compliance judgment of standardized gesture operations. The proposed system reduces dependence on manual rule configuration and improves both the flexibility of rule adaptation in complex scenarios and the real-time performance of compliance detection.

Keywords: Multimodal data fusion; Large language model; Kinematic features; Gesture detection; Compliance perception

0 引言

轨道交通系统具有线路结构复杂、运行组织严密和安全约束强等特点,列车司机作为行车作业的直接执行者,其关键环节中的操作行为直接关系到列车运行安全与运营秩序。在进站、出站、确认站台门等场景中,司机通常需要按照对应的作业制度完成手势动作^[1]。与普通交互手势不同,司机手势具有明确的业务指向,其动作方向、执行顺序和完成时机均与列车运行状态密切相关,因此仅依靠单一视觉特征难以完成准确的合规性判断。本文以轨道交通司机手势合规检测为典型应用场景,开展多模态数据融合与运动学特征驱动的手势动态检测系统研究。

现有相关研究多基于视频图像进行司机手势识别,其

中YOLOv5等目标检测方法^[2]常被用于手部或人体关键区域检测,在目标定位和静态姿态识别方面具有一定优势。然而,这类方法通常更关注图像帧中的目标位置或局部动作特征,对手势连续运动过程、动作方向变化及作业场景语义的刻画相对不足。

为此,本文在基于运动学特征的手势动态检测方法基础上,引入多模态数据融合思想,将路网信息、司机室视频流和列车运行状态数据纳入统一感知框架。系统通过语义解析将路网信息转化为结构化规则数据,并结合当前列车运行状态动态匹配轨迹判定帧数与手势检测规则,从而减少对人工预设规则的依赖,使合规检测过程更符合实际作业逻辑。

1 系统总体架构

手势动态检测系统面向司机作业过程中的动态合规检测需求,由多模态数据接入层、语义解析模块、规则匹配模块、手势运动学检测模块、合规判定与预警模块构成。各模块围绕“场景识别—规则生成—动作检测—合规验证”的闭环进行组织。系统首先接入路网信息、司机室视频和列车运行状态信息,随后将路网侧的非结构化文本转化为结构化规则数据;当系统获得列车当前运行状态后,规则匹配模块根据结构化规则数据,确定当前应采用的轨迹判定帧数和手势检测规则,再将该结果传递给运动学检测模块进行手势序列分析。

与传统单一视频识别方案^[1]相比,该架构的关键差异在于检测规则不再由人工静态配置,而是由路网规则和列车状态共同驱动。系统能够根据进站、出站、启动、停车等不同场景,自动确定检测窗口和手势方向序列,使感知结果具有更明确的业务语义。

1.1 多模态数据接入层设计

多模态数据接入层是系统获取外部信息的基础,主要接入三类数据。第一类为路网信息,覆盖线路、站场、区段和特定场景下的司机确认要求等内容,其中大部分以文本、表格或制度文件形式存在,具有非结构化或半结构化特征。第二类为司机室视频流,用于获取司机在作业过程中的连续图像信息。视频流来自驾驶室摄像头或车载视频采集设备,系统对其进行帧率读取和缓存管理,确保后续能够按照规则匹配结果截取规定数量的连续帧。第三类为列车运行状态数据,包括列车速度、运行方向、站点位置、区间位置、车门状态等。这类数据用于判断列车当前所处的运行场景,是触发规则匹配的主要依据。

为保证合规检测与实际作业场景一致,系统在数据接入阶段对视频流与运行状态进行统一标识和时间对齐。由此,当列车进入特定作业场景时,系统能够准确获取当前应执行的手势确认规则,为后续合规判定提供数据基础。

1.2 语义解析模块

轨道交通路网信息来源复杂,通常包含线路基础信息、作业制度、场景说明及特殊规定等内容^[4],原始形式多为自然语言文本或半结构化表格,难以直接用于程序计算。为提高规则配置效率,本文建立语义解析模块,利用大模型对非结构化路网信息进行语义理解与字段抽取,生成可供系统调用的结构化规则数据。

通过语义解析模块,原本分散在文本和制度文件中的作业要求被转化为机器可理解、可匹配的规则数据。这样

既便于不同线路、站场和车型规则的统一管理,也能在规章调整时通过重新解析与审核完成规则更新,减少人工逐项配置的工作量。

1.3 运行场景规则匹配模块

该模块用于在语义解析模块完成规章解析并生成结构化规则数据后,结合当前列车运行状态信息进行规则匹配,输出当前场景下对应的轨迹判定帧数和手势检测规则。系统首先接收列车运行状态数据,提取速度、停车或启动状态等关键字段,形成当前运行场景的状态描述;随后将这些状态信息与结构化规则数据进行比对,查找与当前线路、车型、站场条件和作业环节相匹配的规则条目,并据此确定当前场景所对应的检测参数。

匹配结果可直接供后续合规检测模块调用,主要包括当前手势方向规则序列、轨迹判定帧数以及需要截取的视频流帧数等内容。通过这种方式,系统能够根据列车实时运行状态和既有规则表动态确定检测参数,避免人工预先固定帧数和规则序列,使合规检测过程能够随不同线路、不同站场及不同作业阶段的变化自适应调整,从而提升规则适配的灵活性与检测流程的场景一致性。

1.4 基于运动学特征的手势动态检测方法

1.4.1 视频帧截取与手部中心点提取

当规则匹配模块输出当前场景所需的帧数后,系统从司机室视频流中截取相应的连续帧序列,截取得到的图像序列帧数为当前场景匹配得到的视频帧数量。系统对每一帧图像进行手部目标检测,可采用目标检测算法^[5]识别手部区域边界框。边界框以像素坐标表示,包含左上角顶点坐标、宽度和高度。根据边界框计算手部中心点坐标,得到每一帧对应的手部中心点序列。形成的手部运动轨迹序列,为后续动态分析提供基础。

1.4.2 手势运动参数计算与转折点定位

在获得连续帧手部中心点坐标后,系统基于相邻两帧中心点位置变化计算手部运动速度,并依据相邻速度之间的变化计算加速度。速度用于判断当前图像序列中是否存在明显手部运动,加速度用于刻画运动状态变化,并辅助确定手部运动的转折点。

转折点表示手部运动状态发生关键变化的帧位置。在连续手势确认过程中,司机通常会从身前位置指向某一目标区域,再回到身前位置,手部速度会经历加速、减速或方向变化过程。系统可将加速度接近零且前后运动趋势发生变化的位置作为候选转折点,并结合中间帧位置进行验证。这样既能避免仅凭单帧位置判断方向造成误差,也能

保留手势动态变化的连续性。

1.4.3 有效动作判断与方向识别

为避免将轻微抖动或无关动作误判为有效手势，系统首先依据最大速度进行初筛。若图像序列中的最大速度低于当前规则配置的速度阈值，则判定该序列不存在有效手势动作，检测流程终止；若最大速度达到阈值要求，则继续进行轨迹幅度判断。系统选取图像序列中间帧，分别计算中间帧手部中心点与序列首帧、末帧之间的距离，若两者均低于当前规则配置的距离阈值，则说明轨迹变化不明显，不满足有效动作要求；若至少存在一个距离超过阈值，则判定该序列存在有效动作。

方向识别阶段，系统对图像中的头部区域进行检测，提取头部中心点的横向坐标作为方向判断基准。随后计算转折点头部横坐标与头部中心点横坐标之间的差值，并引入中间帧差值进行辅助验证。当转折点和中间帧均位于头部中心点右侧且超过右向阈值时，判定手部向右运动；当两者均位于阈值范围内时，判定为向中运动；当两者均位于头部中心点左侧且超过左向阈值时，判定为向左运动。由于转折点只反映某一瞬时状态，中间帧可用于验证方向趋势是否持续，从而提升方向判断的稳健性。

1.4.4 滑动窗口方向序列生成

司机完成一整套手势时，动作通常包含多个方向或多个确认点。为从连续图像序列中提取完整方向序列，系统采用滑动窗口方式进行局部方向判断。窗口长度由规则匹配模块输出，并小于当前场景下一整套手势对应的帧数。窗口从手部中心点序列起始位置开始，按照逐帧递进方式滑动，每次形成一个连续且与前一窗口重叠的局部序列。

每个窗口独立执行有效动作判断和方向识别，得到对应的局部方向结果。随后系统对相邻窗口结果进行筛选：若当前窗口方向与上一窗口方向一致，说明手部仍处于同一方向运动或保持阶段，不重复写入最终序列；若当前窗口方向发生变化，说明手势到达新的方向位置或进入新的确认阶段，将当前方向纳入最终运动方向序列。通过这种方式，系统可以得到一整套手势对应的方向序列，例如“右—左—中”或由具体线路规则定义的其他顺序。

1.5 合规判定与预警机制

合规判定模块接收手势运动学检测模块生成的运动方向序列，并读取运行场景规则匹配模块输出的规则方向序列。两类序列的来源不同：前者来自视频中的实际动作，后者来自当前运行状态和结构化规则数据的匹配结果。系统按照顺序对两类序列进行比对，若方向数量、方向类型

和方向顺序均一致，则判定当前手势符合该作业场景下的规则要求；若存在缺失、顺序错误、方向错误或动作未完成等情况，则判定为违规。

当判定结果为违规时，系统通过车载显示设备显示警示信息，并结合声音提示提醒司机及时纠正。与此同时，系统将违规记录上传至管理端，记录内容包括列车运行场景、触发规则、实际方向序列、标准方向序列、发生时间和相关视频片段索引等。该记录可用于事后复核、司机培训和安全管理分析。

2 系统特点与应用价值

本文系统的技术特点主要体现在三个方面。第一，系统将司机室视频、列车运行状态和路网信息统一纳入合规感知流程，改变了传统方案只依赖单路视频的局限，使手势检测结果能够与真实作业场景相对应。第二，系统通过语义解析模块，将非结构化路网信息转化为可计算规则，提升了不同线路、不同车型和不同站场条件下的规则维护效率。第三，系统运用了基于运动学特征的动态手势检测方法，通过手部中心点轨迹、速度、加速度和转折点分析，能够更准确地描述司机手势的连续变化过程，避免仅凭静态图像进行判断带来的误差。

在应用层面，可服务于列车司机作业行为实时监测、运营安全辅助管理、司机规范化培训和事后追溯分析等场景。与人工抽查或离线视频复核相比，系统能够在违规发生时及时提示；与固定规则检测相比，系统能够随运行场景动态调整检测标准。特别是在多线路并行运营、站场条件差异较大或运营规程频繁更新的情况下，基于结构化规则和场景匹配的合规感知机制具有更好的扩展性。

3 结语

本文以轨道交通司机手势合规检测为典型应用场景，提出了一种多模态数据融合与运动学特征驱动的手势动态检测系统。系统构建了多模态数据接入层，融合路网信息、司机室视频流和列车运行状态数据；设计了语义解析模块，将非结构化路网信息转化为结构化规则数据；并通过运行场景规则匹配模块，将当前列车运行状态与结构化规则进行匹配，动态生成当前场景下的手势检测规则和视频帧截取数量。由此，合规检测所需的图像帧和规则序列均可由场景匹配结果自动确定，无需人工固定预设。

在检测方法上，系统采用手部中心点轨迹分析、速度和加速度计算、转折点定位、有效动作判断、方向识别及滑动窗口序列生成等步骤，实现对司机手势动态过程的合规判定。该系统兼顾业务规则适配与动作识别稳定性，能

能够为轨道交通运行安全管理提供更具场景感知能力的技术支持。后续可进一步围绕多源数据时间同步、规则冲突消解、低照度视频增强以及人工审核闭环等方向展开研究，以提升系统在实际运营环境中的可靠性和可维护性。

参考文献:

[1] 曹海鹏, 唐伟忠, 宋哲超等. 车务段安全生产指挥中心管理系统设计[J]. 铁路计算机应用, 2024, 33(6): 46-56.

[2] 闫颢月, 王伟, 田泽. 复杂环境下基于改进 YOLOv5 的手势识别方法[J]. Journal of Computer Engineering & Applications, 2023, 59(4).

[3] 谢晓燕, 曹盘宇, 夏浩等. 基于编码视频的动态手势数据优化与识别[J]. 北京邮电大学学报, 2024, 47(2): 90.

[4] 杨连报, 李平, 刘艺飞. 文本大数据分析技术在铁路行车安全领域的应用研究[J]. 中国铁路, 2020.

[5] Sun Y, Li Y, Li S, et al. PBA-YOLOv7: an object detection method based on an improved YOLOv7 network[J]. Applied Sciences, 2023, 13(18): 10436.

作者简介: 王乐萌(1999.03-), 女, 汉族, 北京人, 硕士研究生, 助理工程师, 研究方向: 人工智能。